Econometrica Supplementary Material

SUPPLEMENT TO "TEAM PLAYERS: HOW SOCIAL SKILLS IMPROVE TEAM PERFORMANCE"

(*Econometrica*, Vol. 89, No. 6, November 2021, 2637–2657)

BEN WEIDMANN
Harvard Kennedy School, Harvard University

DAVID J. DEMING
Harvard Kennedy School, Harvard University and NBER

APPENDIX: DETAILED TASK DESCRIPTIONS

THIS PART OF THE SUPPLEMENT EXPANDS on Section 2.2 and provides a detailed description of the three tasks we used to estimate the team player effect (Optimization, Memory, and Shapes) along with our validation task (Cryptography). All the tasks aimed to meet three criteria. First, we sought tasks that could be administered to both individuals and groups, with only minor modifications between the individual and group versions. This enabled us to estimate group performance *controlling for individual task-specific skill*. Second, tasks needed to be objective in the sense that we could easily rank performance across individuals and groups. Third, since we are interested in studying teamwork, we looked for tasks where cooperation among group members would plausibly improve performance.

A.1. *Optimization Task*

The goal of this task was to find the maximum of a complex function.[1] Some example functions are presented in Figure A.1 (left panel). In the individual Optimization task, participants were given a function, which was hidden to them, and had 15 guesses to find the maximum. They entered guesses between 0 and 300. For example, a participant attempting to find the maximum of function b (f_b) would see the interface presented in the right panel of Figure A.1. For each guess, the computer returns f_b(guess). Once participants had entered 15 guesses, they were asked to submit their answer for the input value that maximized the output. In Battery A, individuals completed the Optimization task three times. A different underlying function was used each time.

In the group version of the task, each group member was allocated 5 guesses. Collectively, the group had a total of 15 guesses. Each group member entered their own guesses on their own laptop. A critical feature of this task was the need to involve all three group members. After the group had entered its 15 guesses, the Reporter was asked to enter the group's answer for the output-maximizing input. Each group solved the Optimization task twice. Every time participants attempted the Optimization task, they engaged with a new underlying function. Success on the group Optimization task required collective planning and the sharing of unique information. Both these factors have been shown in previous small-group research to predict group performance across a range of contexts (Driskell, Salas, and Driskell (2018), Mesmer-Magnus and DeChurch (2009), Weingart (1992)).

Ben Weidmann: benweidmann@g.harvard.edu
David J. Deming: david_deming@harvard.edu

[1]We developed the Optimization task specifically for the purposes of this experiment. We were inspired by Mason, Jones, and Goldstone (2008), who used a numerical optimization task to study how innovations propagate across networks. The individual task was piloted in a MTurk sample.

 https://doi.org/10.3982/ECTA18461

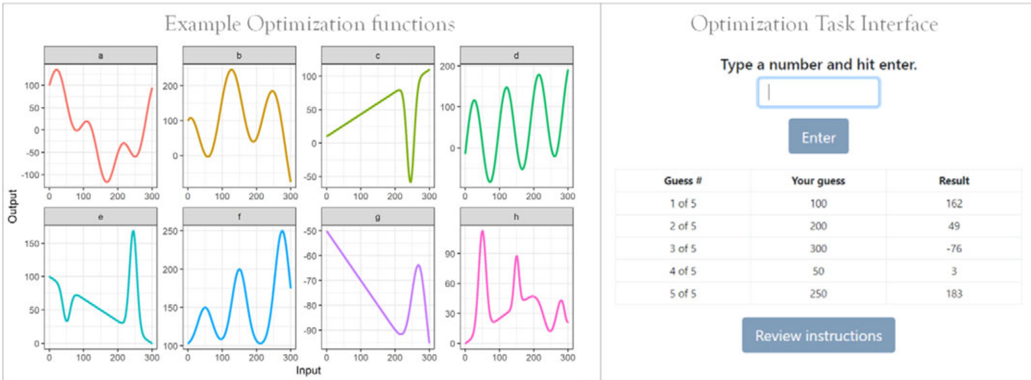


FIGURE A.1.—Description of the Optimization task. *Note*: This figure presents a visual description of the Optimization task. Participants were asked to enter guesses between 0 and 300 (the horizontal axis). They received an output from a complex function (the vertical axis). The left panel contains example functions, which were hidden to participants. The right panel is the participant interface. The goal of the task is to find the maximum of the function. In the first phase of the experiment, individuals received 15 guesses before submitting an answer. In the second phase, each member of a 3-person team received 5 guesses; once all these guesses had been entered, the group agreed upon a final answer.

A.2. *Memory Task*

This task focused on short-term memory, which is closely associated with fluid intelligence and IQ (Colom, Rebollo, Abad, and Shih (2006), Nisbett et al. (2012)). We tested participants' ability to memorize three different types of stimuli: words, images, and stories.[2]

In Phase 1 of the experiment, individuals' short-term memory for each type of stimuli was measured sequentially. Participants began by completing the words test. This involved memorizing a list of 12 target words over 24 seconds (the stimuli come from the Hopkins Verbal Learning Test, reported in Brandt (1991). After the memorization period, participants were presented with sets of three words and were asked to identify which, if any of the three, were target words. Next, participants completed the images test, in which they were given 20 seconds to memorize six target faces (the stimuli come from the Cambridge Face Memory Test, described in Duchaine and Nakayama (2006)). Participants were then presented with 15 sets of three faces and asked to identify target faces. Last, participants completed the stories test in which they had 40 seconds to read two short paragraphs, of roughly 60 words each. The stimuli were adapted from Wechsler Logical Memory III (Wechsler (1997)). At the end of the memorization period, participants were asked nine multiple choice questions about the two paragraphs.

Once participants had completed the three individual memory tests, we provided feedback about their results. This included information on an individual's overall performance relative to other participants and emphasized the test on which they scored highest. Our

[2]We drew on a model of memory that emphasizes three subsystems: verbal, visual-spatial, and episodic (Baddeley (2001)). Our three stimuli map onto these subsystems: verbal → words; visualspatial → images; episodic → stories. We note that the Baddeley model focuses on working memory, not short-term memory. The two concepts, however, are very closely linked, as discussed in Colom et al. (2006). The reason we focus on short-term memory is that the subtests are easier to translate into a practical task for groups to perform when working face-to-face in a lab setting.

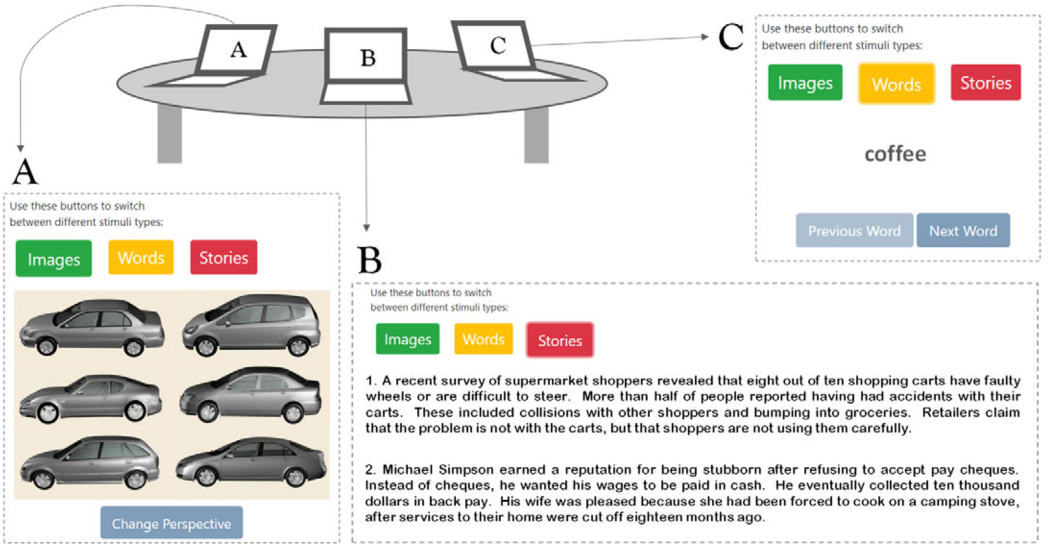


FIGURE A.2.—Description of the Memory task (for groups). *Note*: This figure presents a visual description of the group Memory task. Participants were given 40 seconds in which they could cycle through: 6 different images (Panel A), 2 different stories (Panel B), and 12 different words (Panel C) on their own laptop. Then groups gathered around a single laptop and answered 24 questions together about the three sets of stimuli, with an equal number of questions about each.

goal with the feedback was to provide people with information they might use in the group phase of the experiment to select sub-tasks on which they were most proficient.

In the group version of the task, we combined established measures of individual memory into a collaborative memory challenge. Each group was given 40 seconds to collectively remember 12 words, 6 images, and 2 stories. We added story and images stimuli to those described above, so that each time a group encountered the Memory task they were asked to memorize unseen material.[3]

Each member of the group viewed their own laptop and could view any of the three stimuli. Participants could change the stimuli they were memorizing during the 40-second memorization period. In the example presented in Figure A.2, participant A is memorizing images (cars), participant B is memorizing stories, and participant C is memorizing words. During the 40-second memorization period, participants could change the stimulus they were viewing at any time by using the buttons in the top left of their screens. Before the memorization period began, groups were prompted to discuss their strategy.

After the memorization period, all three team members gathered around the Reporter's laptop to answer a set of 24 questions about the stimuli. There were an equal number of questions about each type of stimuli. The structure of the questions mirrored those used in the individual assessments.

A.3. *Shapes Task*

This task relied on two well-established measures of fluid intelligence: the Culture Fair Intelligence Test (CFIT, Scale 3) and the Raven's Advanced Progressive Matrices

[3]We supplemented stories with shortened versions from Sullivan (2005). For images, we added related tests focused on cars, bikes, and bodies, described in Dennett et al. (2012).

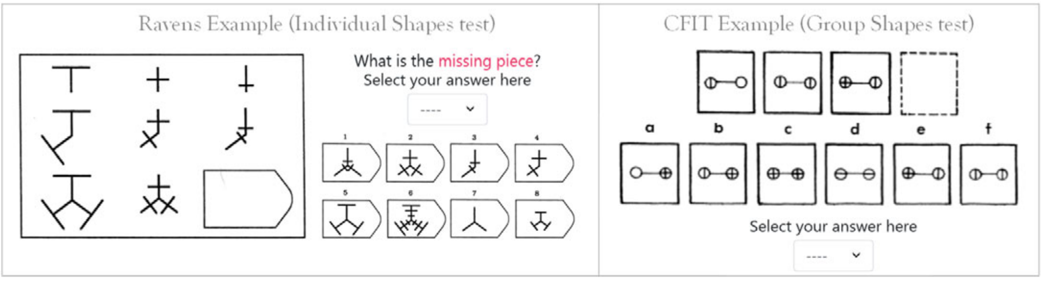


FIGURE A.3.—Description of the Shapes task. *Note*: This figure presents example items from the Shapes task, which was adapted from two well-established measures of IQ or fluid intelligence—the Culture Fair Intelligence Test (CFIT) and the Raven's Advanced Progressive Matrices (Ravens). In the individual phase of the experiment, participants were given 14 items and seven minutes. The mean score was 7.3, and no individual received a perfect score. In the group phase, all members gathered around a single laptop and collectively decided on an answer for each item. Mean scores differed by battery but once again no groups received a perfect score. The correct answer for the Ravens Example is '1'; for the CFIT example the correct answer is 'c'.

(Ravens). In the individual testing phase, participants completed 14 Ravens items (even numbered items, ranging in difficulty from across sets I and II; see Raven (2003)).

This task centers on pattern recognition and spatial reasoning. Participants are asked to look for a pattern and determine "what comes next." As an example, consider the pattern established in the left-most box of Figure A.3, which is missing a piece. Participants were asked to find the missing element (from options a to f).

The group version of this task employed the CFIT, which is very similar to the Ravens task. All group members gathered around the Reporter's laptop and collectively decided on the group's answer for each item. In each battery is contained a different form of the CFIT. An example item is provided in the right-hand panel of Figure A.3.

A.4. *Validation Task: Cryptography*

We use the three tasks above to estimate individual contributions to group performance, as described in Section 3. We chose the Cryptography task as a fourth, out-of-sample validation measure of group performance. The Cryptography task is a decoding problem in which each letter from A to J represents a unique number from 0 to 9. Groups were asked to decode the value of each letter by entering mathematical expressions that would return an output (e.g., if A = 5, B = 1, C = 4, and D = 0, an entry of A + B + C would return the value "BD," for 10). An example is shown in Figure A.4.

The procedure for decoding each letter is somewhat complex and is well described elsewhere (Larson (2010)).[4] The goal of the task is to find the value of each letter in the fewest number of steps. We administered this task twice: once as a practice, to make sure that groups understood the process—and a second time to assess their performance. Cryptography is one of the very few established tasks that demonstrates "strong synergy"

[4]In brief, the process involved three steps. *Step 1: enter an 'equation'*. An equation is a set of letters with '+' or '–' operators; for example, A + B + C. The computer then returned the answer. If A = 3, B = 1, C = 2, D = 6, then the computer would reply A + B + C = D. *Step 2: make a hypothesis*. Here, a group might guess that D was a large number (as it's the sum of 3 numbers). So, they might guess "D = 7." The computer would reply "FALSE." *Step 3: guess all the values*. The group is allowed, but not compelled, to submit a value of each letter. If all their guesses are correct, the task ends. If not, the group goes back to step 1.

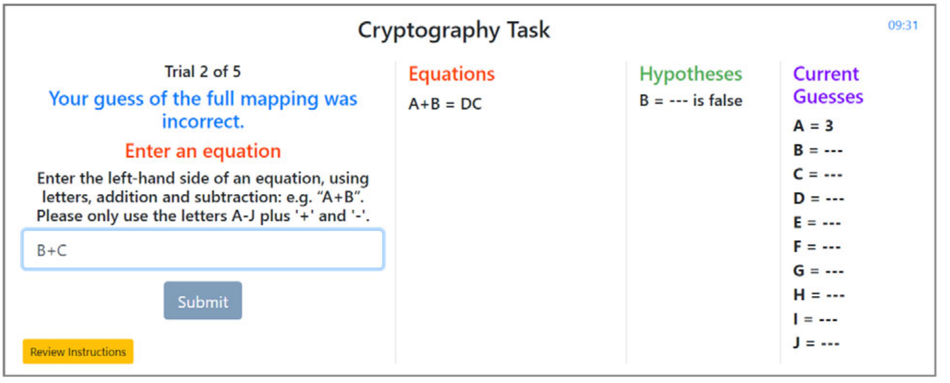


FIGURE A.4.—Description of the Cryptography task. *Note*: This figure presents the Cryptography task interface. Cryptography was a fourth, out-of-sample validation task that was not used to estimate the team player index. Each letter from A to J represents a unique number from 0 to 9. Groups were asked to decode the value of each letter by entering mathematical equations that would return an output. The goal was to decode the letters using as few equations as possible. Groups were given one practice try on the Cryptography task, to make sure they understood the task. Each group was allowed up to 15 equations on the assessed version of the task. Those who decoded all the letters (81%) used a mean of 7.9 equations. Eighty-five groups attempted Cryptography.

in the sense that groups perform better than the sum of their parts (Larson (2010)).[5] This task was only administered in Battery E, the last set of group tasks.

In the main analysis, we estimate the team player effect (σ_β) using a multilevel model. This assumes that the distribution of β is normal. Here we assess that assumption by comparing the fixed effect estimates of $\hat{\beta}_i$ to a normal distribution. The distributions are closely matched.

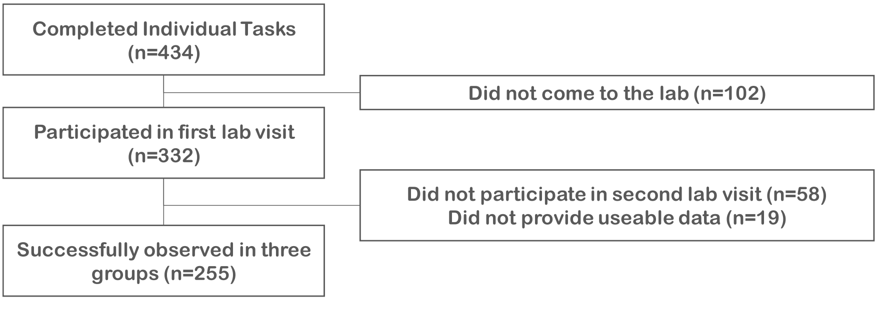


FIGURE A.5.—Participant flow diagram. *Note*: This figure presents the participant flow for the experiment. For details, see Section 2.3.

[5]The reason may be that the task naturally lends itself to people taking on different roles. While some people are figuring out what the next equation should be according to the current strategy, others can consider better strategies. This gives groups the potential to be strategically flexible. Individuals, on the other hand, find it extremely challenging to simultaneously execute a strategy *and* to consider a new one, perhaps due to constraints of attention and working memory (Larson (2010, p. 154)). See also Laughlin, Hatch, Silver, and Boh (2006) and Laughlin, Bonner, and Miner (2002). Note that the underlying feature of the task that enables differentiation is that it is possible to switch strategies at any point in the task, without incurring a cost.

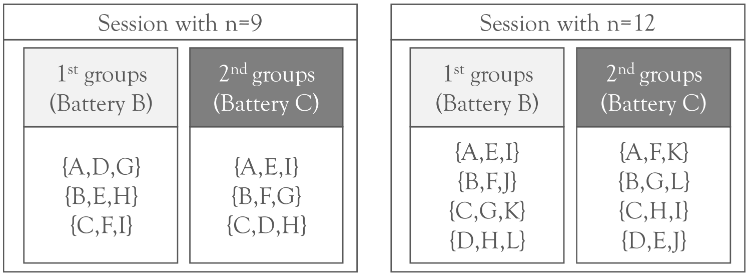


FIGURE A.6.—Example of blocked randomization scheme for first lab visit. *Note*: This figure presents a visual example of how individuals were randomized to groups over the course of a single lab visit. We use Lab Visit One as an example. The left panel illustrates the randomization process for a session of 9 people; the right panel is the equivalent process for a session of 12 people. Participants were randomized to two successive groups in a single draw, and the randomization was blocked so that, where possible, participants did not have any of the same team members in their second group assignment of a lab session. See Section 2.4 for details.

As noted in Section 3, another approach to identifying team players is to estimate (model 1b). Using variables defined in Section 3, we have

$$\tilde{G}_{gk} = \alpha_k \Sigma_i I_g^i X_{ik} + \Sigma_i I_g^i \beta_i + \epsilon_{gk}. \qquad \text{(model 1b)}$$

This approach yields very similar results to our preferred identification strategy. This is illustrated by Figure A.8, which is analogous to Figure 3 in the paper but uses (model 1b) to estimate β fixed effects. Also see Table A.I for more on the robustness of our results to a single-step estimation strategy.

Our pre-analysis plan for model (1) included two indicator variables measuring "group familiarity." These were indicators for whether group g contained participants who knew each other from outside the experiment (5 percent of the sample) and for whether groups contained participants who had previously been assigned to the same team by chance (41 percent of the sample). Neither of these nuisance controls has any substantive impact on the main results—as illustrated in Column 2 of Table A.I—so we dropped them for clarity. Table A.I also illustrates our core result using an alternative, single-step estimation approach. This approach is mentioned in Section 3. Using variables defined in that

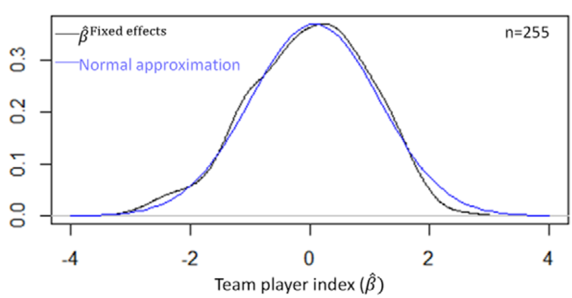


FIGURE A.7.—Distribution of team player index.

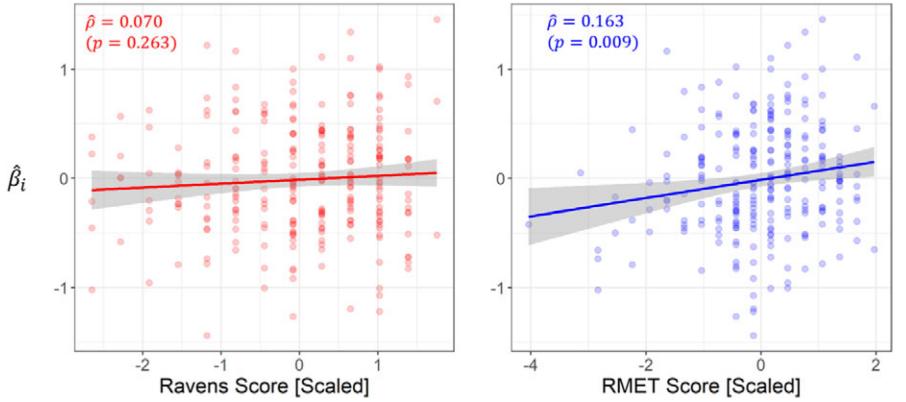
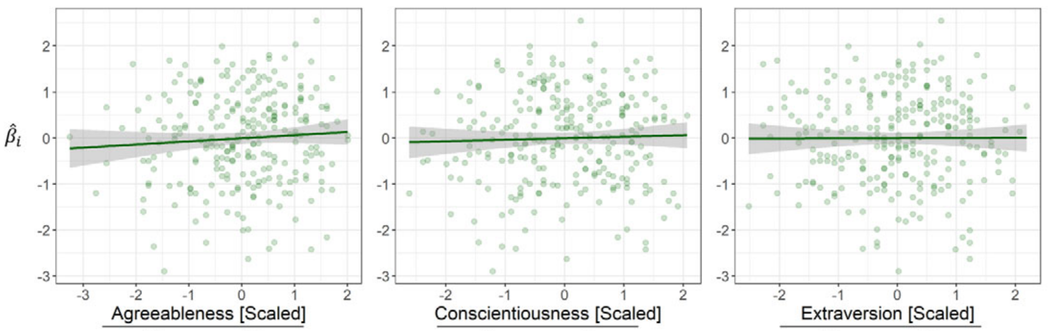


FIGURE A.8.—Correlation between Team Player Effect and IQ and RMET, simultaneous estimation. *Note*: This figure is analogous to Figure 3 in the paper. Here, however, $\hat{\beta}$ has been estimated using the single step approach (model 1b). Each panel of the figure presents a scatterplot of an individual's estimated team player index $\hat{\beta}_i$ against their individual Ravens score (left panel) and their individual RMET score (right panel). Ravens is a well-established measure of IQ or fluid intelligence. RMET is the Reading the Mind in the Eyes Test, a well-established test of emotion perception and social intelligence. The same sample was used for all analysis: 1029 group-task observations, 343 groups, 255 participants.

section, we estimate:

$$\tilde{G}_{igk} = \alpha_k \Sigma_i I_g^i X_{ik} + \beta_i + \epsilon_{igk},$$
$$\beta_i \sim N(0, \sigma_\beta^2), \qquad \text{(model 2b)}$$
$$\epsilon_{igk} \sim N(0, \sigma_g^2).$$



	Agreeableness (1)	Agreeableness (2)	Conscientiousness (1)	Conscientiousness (2)	Extraversion (1)	Extraversion (2)
	0.068	0.031	0.033	0.010	0.004	0.012
	(p=0.279)	(p=0.618)	(p=0.597)	(p=0.879)	(p=0.953)	(p=0.849)

FIGURE A.9.—The team player index is uncorrelated with personality scores. *Note*: Each panel presents a scatterplot of an individual's estimated team player index $\hat{\beta}_i$ against their individual scores on the Agreeableness (left panel), Conscientiousness (middle panel), and Extraversion (right panel) scales of the Big 5 Personality inventory. In all three cases the $\hat{\beta}_i$ shown in the figures is estimated based on the model in equations (1) through (4), as described in Section 3 and detailed in our pre-registered analysis plan. Beneath each panel, we show coefficients from two different estimates of $\hat{\beta}_i$: (1) our pre-specified model, with controls for task-specific skills and indicators for group familiarity; (2) no controls. See the text for details. The scatterplot always shows estimates from model (1). The same sample was used for all analysis: 1029 group-task observations, 343 groups, 255 participants.

TABLE A.I

ROBUSTNESS OF TEAM PLAYER EFFECT TO GROUP FAMILIARITY CONTROLS, AND SINGLE-STEP ESTIMATION.

	Dependent Variable: Group Performance $\tilde{G}_{gk}$		
	(1)	(2)	(3)
Teamplayer Effect $\hat{\sigma}_\beta$	0.129	0.127	0.139
(randomization inference)	$(p = 0.026)$	$(p = 0.029)$	$(p = 0.038)$
[profile likelihood]	$[p = 0.034]$	$[p = 0.037]$	$[p = 0.005]$
{normal approximation}	$\{p < 0.001\}$	$\{p < 0.001\}$	$\{p \leq 0.001\}$
Task-specific skills			
Memory°	0.166	0.166	0.160
	(0.032)	(0.032)	(0.018)
Optimization°	0.125	0.125	0.119
	(0.031)	(0.031)	(0.018)
Ravens (Shapes)°	0.302	0.302	0.299
	(0.030)	(0.030)	(0.017)
Group familiarity controls?		✓	
Alternative analysis (model 2b)			✓
Number of groups	343	343	343
Number of participants	255	255	255

Note: Column 1 reproduces the analysis presented in Table II of the main paper. Column 2 illustrates the effect of the group familiarity controls. Column 3 illustrates the effect of a single-step estimation approach (model 2b). °Indicates group-level sum. "Task-specific skills" means that $\tilde{G}_{gk}$ is conditioned on the mean performance of group g's individuals on task k, that is, $\overline{X}_{gk} = \frac{1}{3}\sum_i I_g^i X_{gk}$. Covariate coefficients have standard errors in parentheses. Estimates of the Teamplayer Effect ($\hat{\sigma}_\beta$) have p-values from randomization inference in parentheses, profile likelihood p-values in brackets, and p-values from a Wald test in braces; p-values are from a null hypothesis test that $\sigma_\beta = 0$.

The results are presented in Column 3 of Table A.I.

Our experiment can be thought of as three mini experiments, one for each task. As per the pre-analysis plan, we report analyses by task type. For each task k (Optimization, Memory, and Shapes), we generated task-specific estimates of our measure $\hat{\beta}_i^k$ and the team player effect $\hat{\sigma}_\beta^k$ using the same analytic approach outlined in Section 3. Table A.V reports on three core parameters.

The task in which individual skills most easily translate to group success is Shapes. This is as expected. First, the underlying Shapes tests have benefited from many years of measurement development and are less noisy than our novel Memory and Optimization tasks. Second, the Shapes task arguably had the least cooperation requirement. On this point, we note that the Shapes task had the lowest point estimate for the team player effect across the three tasks ($\hat{\sigma}_\beta^{\text{Shapes}} = 0.19$). Of the two other tasks, Memory exhibited larger team player effects ($\hat{\sigma}_\beta^{\text{Memory}} = 0.28$ compared to $\hat{\sigma}_\beta^{\text{Optimization}} = 0.24$). The team player index measured using the Memory task also has a strong and statistically significant association with individual RMET ($\hat{\rho} = 0.15$, $p = 0.02$).

In Table A.VI, we make use of $\hat{T}_g$ estimates in an effort to understand the characteristics associated with group "efficiency"—defined as "group performance, after controlling for individual skill." We explore the association between $\hat{T}_g$ and the following eight group-

TABLE A.II

TEAM PLAYER EFFECT WITH MANY FUNCTIONS OF INDIVIDUAL SKILL.

	Dependent Variable: Group Performance $\tilde{G}_{gk}$									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Teamplayer Effect $\hat{\sigma}_\beta$	0.129	0.202	0.155	0.131	0.129	0.209	0.184	0.211	0.204	0.120
	$(p = 0.026)$	$(p < 0.001)$	$(p = 0.006)$	$(p = 0.023)$	$(p = 0.026)$	$(p < 0.001)$	$(p < 0.001)$	$(p < 0.001)$	$(p < 0.001)$	$(p = 0.044)$
Task-specific skills										
Sum ($\sum X_{ki}$)	✓									✓
Max ($\max_g X_{ki}$)		✓								✓
Min ($\min_g X_{ki}$)			✓							✓
All (X_{k1}, X_{k2}, X_{k3})				✓						✓
Flexible ($f_k(\sum X_{ki})$)					✓					
Ravens										
Max ($\max_g$ Ravens$_i$)						✓				✓
Min ($\min_g$ Ravens$_i$)							✓			✓
RMET										
Max ($\max_g$ RMET$_i$)								✓		✓
Min ($\min_g$ RMET$_i$)									✓	✓

Note: The team player effect is the standard deviation of our estimate of the causal contribution of each individual to group performance—see Section 3 for details. Estimates of the team player effect ($\hat{\sigma}_\beta$) have p-values in parentheses from the randomization inference procedure. “Flexible” controls for task-specific skills were implemented using splines in a generalized additive model (for equation (1)). Ravens is a well-established measure of IQ or fluid intelligence. RMET is the Reading the Mind in the Eyes Test, a well-established test of emotion perception and social intelligence. Group familiarity controls are indicators for whether any members of the group knew each other prior to randomization, and whether any members of the team had been on a team together during a previous lab visit. The same sample was used for all analysis: 1029 group-task observations, 343 groups, 255 participants.

TABLE A.III

WHAT IS THE MOST PREDICTIVE MEASURE OF INDIVIDUAL SKILL?

Dependent Variable: Group Performance on *Shapes* task $\tilde{G}_{g\,\mathrm{Shapes}}$				
Shapes ($\sum_i I_g^i X_{i,\mathrm{shapes}}$)	0.302			
	(0.027)			
Memory ($\sum_i I_g^i X_{i,\mathrm{Memory}}$)		0.193		0.165
		(0.032)		(0.032)
Optimization ($\sum_i I_g^i X_{i,\mathrm{Opt}}$)			0.163	0.131
			(0.032)	(0.031)
Adj R^2	0.262	0.092	0.070	0.134
Number of groups	343	343	343	343
Dependent Variable: Group Performance on *Optimization* task $\tilde{G}_{g\,\mathrm{Opt}}$				
Optimization ($\sum_i I_g^i X_{i,\mathrm{opt}}$)	0.125			
	(0.032)			
Memory ($\sum_i I_g^i X_{i,\mathrm{Memory}}$)		0.036		0.018
		(0.034)		(0.037)
Shapes ($\sum_i I_g^i X_{i,\mathrm{shapes}}$)			0.051	0.045
			(0.032)	(0.035)
Adj R^2	0.040	0.004	0.005	0.002
Number of groups	343	343	343	343
Dependent Variable: Group Performance on *Memory* task $\tilde{G}_{g\,\mathrm{Memory}}$				
Memory ($\sum_i I_g^i X_{i,\mathrm{memory}}$)	0.166			
	(0.033)			
Optimization ($\sum_i I_g^i X_{i,\mathrm{opt}}$)		0.096		0.042
		(0.032)		(0.036)
Shapes ($\sum_i I_g^i X_{i,\mathrm{shapes}}$)			0.130	0.111
			(0.031)	(0.035)
Adj R^2	0.067	0.023	0.046	0.047
Number of groups	343	343	343	343

level characteristics:[6]

1. XSUM_g: the overall level of individual skill in group g (defined as $\sum_i I_g^i \overline{X}_i$)
2. XMax_g: the individual skill level of the strongest member of group g ($\max(\overline{X}_i)$ in group g)
3. $\mathrm{RMET\,SUM}_g$: sum of individual RMET in group g ($\sum_i I_g^i \mathrm{RMET}_i$)
4. DivALL_g = diversity of each group, with respect to age; gender; education; and individual skill[7]
5. DivX_g: diversity of skill in group g ($\mathrm{Max}_g - \mathrm{Min}_g$)
6. DivETH_g: binary indicator = 1 if all group members identify as the same ethnicity

[6]For these analyzes, $\overline{X}_i$ is defined as the average score for participant i across Memory, Shapes, and Optimization. Each of these task scores is scaled and averaged: $\overline{X}_i = \frac{1}{3}\sum_k \tilde{X}_{ik}$, where $\tilde{X}_{ik} = \frac{X_{ik} - \hat{\overline{X}}_k}{\hat{s}_k}$, and $\hat{s}_k$ is the sample standard deviation of X_{ik}.

[7]The DIV_g variable summarizes group diversity. Specifically, it is the mean Mahalanobis distance between all group members. If group g has members a, b, c, this is given by $\mathrm{DIV}_g = \frac{1}{3}(d_{ab} + d_{ac} + d_{bc})$, where d_{ab} is the Mahalanobis distance between a and b defined with respect to X_i, gender_i, $\mathrm{education}_i$, and age_i.

TABLE A.IV

EXPLORING MEASUREMENT ERROR WITH SIMULATION.

Level of Measurement Error	$\overline{\hat{\sigma}}_\beta$ Estimated Teamplayer Effect	$\overline{\hat{\rho}}_{\hat{\beta},\text{RMET}}$ Correlation Between $\hat{\beta}_i$ and RMET_i
$c = 0$	0.107	0.111
	$(p = 0.087)$	$(p = 0.074)$
$c = 0.25$	0.114	0.121
	$(p = 0.072)$	$(p = 0.057)$
$c = 0.50$	0.131	0.141
	$(p = 0.038)$	$(p = 0.029)$
$c = 0.75$	0.152	0.169
	$(p = 0.015)$	$(p = 0.011)$
Original Results	0.129	0.156
(for comparison)	$(p = 0.026)$	$(p = 0.012)$

Note: This table presents the results of simulations in which measurement error is added to X_{ik} (measures of individual task-specific skill, assessed in Phase 1 of the experiment). The first four rows of the table use (model 1c) to estimate the team player effect. This model, which is not our pre-registered or preferred approach, implicitly assumes that individual contributions to group performance are determined by a single latent construct that is noisily measured by each individual task. Each draw in the simulation $j = 1, \dots, 500$ follows three steps. First, we add an amount of noise c to X_{ik} (which are initially scaled to have a mean of zero and a standard deviation equal to 1). This gives us a new measure of individual skill $X_{ik}^j \sim N(X_{ik}, c^2)$. To make X_{ik}^j comparable to X_{ik}, we scale it so that $\text{var}(X_{ik}^j) = 1$. Second, we estimate (model 1c), using X_{ik}^j instead of X_{ik}. Third, we proceed with the analysis described in Section 3, and estimate models 2, 3, and 4. We then report $\hat{\sigma}_\beta^j$ and $\hat{\rho}^j_{\hat{\beta}_i,\text{RMET}_i}$. The numbers reported in Table A.IV are the mean values across our 500 simulations, $\overline{\hat{\sigma}}_\beta = \frac{1}{500}\sum_j \hat{\sigma}_\beta^j$ and $\overline{\hat{\rho}}_{\hat{\beta}_i,\text{RMET}_i} = \frac{1}{500}\sum_j \hat{\rho}^j_{\hat{\beta}_i,\text{RMET}_i}$. Similarly, p-values are means across the 500 simulations. The p-values for $\hat{\sigma}_\beta^j$ tests the null that $\hat{\sigma}_\beta^j = 0$ using randomization inference.

7. DivGEN_g: binary indicator = 1 if all group members identify as the same gender
8. DivEG_g: binary indicator = 1 if all group members are the same gender and ethnicity

Table A.VI lists the correlations between various characteristics and $\hat{T}_g$.

Teams with a greater endowment of emotional perceptiveness perform above average in terms of group efficiency (T_g). In other words, teams with higher RMET scores were

TABLE A.V

IS THE TEAM PLAYER INDEX SENSITIVE TO TASK TYPE?

	Analysis by task-type		
	$\hat{\sigma}_\beta^k$	$\rho^k_{\hat{\beta},\text{RMET}}$	$\rho^k_{X,G}$
Memory	0.28	0.14	0.26
	$(p = 0.001)$	$(p = 0.022)$	$(p < 0.001)$
Shapes	0.19	0.15	0.51
	$(p = 0.046)$	$(p = 0.017)$	$(p < 0.001)$
Optimization	0.24	−0.01	0.21
	$(p = 0.018)$	$(p = 0.885)$	$(p < 0.001)$

Note: This table presents core results when each task was analyzed separately. We report 3 parameters: $\hat{\sigma}_\beta^k$ is the estimated team player effect for task-type k; $\rho^k_{\hat{\beta},\text{RMET}}$ is the correlation between $\hat{\beta}_i$ measure for task-type k, and RMET_i; and $\rho^k_{X,G}$ is the estimated correlation between group scores G_{gk}, and each group's endowment of individual skill on task k. $n = 255$ individuals and 343 groups for all cells. p-values for $\hat{\sigma}_\beta^k$ come from randomization inference.

TABLE A.VI

WHAT ARE THE CHARACTERISTICS OF TEAMS THAT ARE "MORE THAN THE SUM OF THEIR PARTS"?

	Correlation With $\hat{T}_g$	p-Value	Number of Groups
Group skills variable			
1. Xsum	0.074	0.169	343
2. Xmax	0.006	0.913	343
3. RMETsum	0.174	0.001	343
Group diversity variable			
4. DivALL	−0.017	0.756	342
5. DivX	−0.056	0.299	343
6. DivETH	0.033	0.541	341
7. DivGEN	−0.005	0.933	332
8. DivEG	−0.039	0.470	332

better at translating individual skills into group performance. Equally, in exploratory analyses, we find that the minimum value of individual RMET in a group (RMET_MIN$_g$) is associated with T_g ($\hat{\rho} = 0.16$, $p = 0.004$), suggesting that having one person who struggles on RMET is enough to limit a team's ability to translate its skills into outputs. None of the diversity variables are significantly associated with skill-adjusted group scores.

In accordance with our analysis plan, we checked whether there was an association between the team player index and "Reporter" status. Most participants (164 out of 255) were the Reporter in one or two of their four groups. Six people were the Reporter in all four of their groups.

We found no evidence of an association between Reporter status and the team player index. The correlation between the number of times a participant was a Reporter and the team player index is $\hat{\rho} = -0.01$ ($p = 0.85$). The mean team player index did not vary by the number of times a participant was the Reporter.

TABLE A.VII

IS BEING THE "REPORTER" ASSOCIATED WITH THE TEAM PLAYER INDEX?

	# of Times Being the Reporter (Out of 4)				
	0	1	2	3	4
# of participants	64	99	65	21	6
Mean team player index ($\overline{\hat{\beta}}$)	−0.020	0.006	0.068	−0.129	−0.170

Note: In this table, $\hat{\beta}_i$ are scaled so that mean$(\hat{\beta}_i) = 0$ and sd$(\hat{\beta}_i) = 1$.

TABLE A.VIII

RELATIONSHIP BETWEEN RMET AND THE TEAM PLAYER INDEX.

	Dependent Variable: $\hat{\beta}_i$ (Pre-Specified Model)			
	(1)	(2)	(3)	(4)
RMET	0.156	0.158	0.182	
	(0.062)	(0.064)	(0.068)	
Age		−0.074	−0.117	−0.094
		(0.064)	(0.073)	(0.074)
Female		−0.019	−0.031	−0.010
		(0.064)	(0.065)	(0.065)
Years of education		0.034	0.038	0.013
		(0.090)	(0.091)	(0.092)
Years of education^2		−0.037	−0.031	−0.041
		(0.089)	(0.090)	(0.091)
Ravens			−0.072	−0.007
			(0.076)	(0.072)
Personality				
Agreeableness			0.059	0.070
			(0.070)	(0.071)
Extraversion			−0.002	−0.007
			(0.068)	(0.069)
Conscientiousness			0.041	0.023
			(0.067)	(0.068)
Observations	255	250	250	250
R^2	0.024	0.033	0.043	0.014
Adjusted R^2	0.021	0.013	0.007	−0.018

Note: Each column presents a regression in which the dependent variable is the team player index ($\beta\beta ii$) from our pre-registered model described in Section 3. RMET is the Reading the Mind in the Eyes Test, a well-established test of emotional perception and social intelligence. Ravens is a well-established measure of IQ or fluid intelligence. Personality comes from three of the five factors in the “Big 5” personality inventory. Covariate coefficients have standard errors in parentheses. All variables were standardized to have mean = 0 and sd = 1. The same sample was used for all analysis: 1029 group-task observations, 343 groups, 255 participants.

TABLE A.IX

CORRELATION BETWEEN INTELLIGENCE AND TEAM PLAYER EFFECT—INTERACTION TERMS.

	Dependent Variable: $\hat{\beta}_i$ (No Controls)				
	(1)	(2)	(3)	(4)	(5)
RMET	0.300	0.199	0.220	0.312	0.233
	(0.060)	(0.061)	(0.064)	(0.061)	(0.063)
Ravens		0.308	0.230		0.233
		(0.061)	(0.070)		(0.069)
Demographics					
Age			−0.138		−0.138
			(0.067)		(0.067)
Female			−0.013	0.003	−0.014
			(0.060)	(0.061)	(0.060)
Years of education			−0.009		−0.009
			(0.084)		(0.084)
Years of education^2			−0.041		−0.038
			(0.082)		(0.082)
Personality					
Agreeableness			0.023		0.027
			(0.065)		(0.065)
Extraversion			0.028		0.026
			(0.063)		(0.063)
Conscientiousness			0.068		0.067
			(0.062)		(0.062)
Interactions					
RMETRavens		−0.027	−0.033		
		(0.059)	(0.059)		
RMETFemale				0.055	0.031
				(0.060)	(0.058)
Observations	255	255	250	252	250
R^2	0.090	0.179	0.192	0.095	0.192
Adjusted R^2	0.086	0.169	0.158	0.084	0.158

Note: Each column presents results from a regression of an individual's team player index $\hat{\beta}_i$, estimated from a model in which we do not control for individual task-specific skills (i.e., the results presented in Column 2 of Table II). RMET is the Reading the Mind in the Eyes Test, a well-established test of emotion perception and social intelligence. Ravens is a well-established measure of IQ or fluid intelligence. Personality comes from three of the five factors in the "Big 5" personality inventory. Task-specific skill is the standardized average of individual scores on all three problem-solving tasks—see the text for details. Covariate coefficients have standard errors in parentheses.

REFERENCES

BADDELEY, A. D. (2001): "Is Working Memory Still Working?" *American Psychologist*, 56 (11), 851–864. [2]

BRANDT, J. (1991): "The Hopkins Verbal Learning Test: Development of a New Memory Test With Six Equivalent Forms," *The Clinical Neuropsychologist*, 5 (2), 125–142. [2]

COLOM, R., I. REBOLLO, F. J. ABAD, AND P. C. SHIH (2006): "Complex Span Tasks, Simple Span Tasks, and Cognitive Abilities: A Reanalysis of Key Studies," *Memory & Cognition*, 34 (1), 158–171. [2]

DENNETT, H. W., E. MCKONE, R. TAVASHMI, A. HALL, M. PIDCOCK, M. EDWARDS, AND B. DUCHAINE (2012): "The Cambridge Car Memory Test: A Task Matched in Format to the Cambridge Face Memory Test, With Norms, Reliability, Sex Differences, Dissociations From Face Memory, and Expertise Effects," *Behavior Research Methods*, 44 (2), 587–605. [3]

DRISKELL, J. E., E. SALAS, AND T. DRISKELL (2018): "Foundations of Teamwork and Collaboration," *American Psychologist*, 73 (4), 334–348. [1]

DUCHAINE, B., AND K. NAKAYAMA (2006): "The Cambridge Face Memory Test: Results for Neurologically Intact Individuals and an Investigation of Its Validity Using Inverted Face Stimuli and Prosopagnosic Participants," *Neuropsychologia*, 44 (4), 576–585. [2]

LARSON, J. R. (2010): *In Search of Synergy in Small Group Performance*. Psychology Press. [4,5]

LAUGHLIN, P. R., B. L. BONNER, AND A. G. MINER (2002): "Groups Perform Better Than the Best Individuals on Letters-to-Numbers Problems," *Organizational Behavior and Human Decision Processes*, 88 (2), 605–620. [5]

LAUGHLIN, P. R., E. C. HATCH, J. S. SILVER, AND L. BOH (2006): "Groups Perform Better Than the Best Individuals on Letters-to-Numbers Problems: Effects of Group Size," *Journal of Personality and Social Psychology*, 90 (4), 644–651. [5]

MASON, W. A., A. JONES, AND R. L. GOLDSTONE (2008): "Propagation of Innovations in Networked Groups," *Journal of Experimental Psychology: General*, 137 (3), 422–433. [1]

MESMER-MAGNUS, J. R., AND L. A. DECHURCH (2009): "Information Sharing and Team Performance: A Meta-Analysis," *Journal of Applied Psychology*, 94 (2), 535–546. [1]

NISBETT, R. E., J. ARONSON, C. BLAIR, W. DICKENS, J. FLYNN, D. F. HALPERN, AND E. TURKHEIMER (2012): "Intelligence: New Findings and Theoretical Developments," *American Psychologist*, 67 (2), 130–159. [2]

RAVEN, J. (2003): *Raven's Advanced Progressive Matrices, Test Books 1 and 2: Pearson Assessment.* [4]

SULLIVAN, K. (2005): "Alternate Forms of Prose Passages for the Assessment of Auditory-Verbal Memory," *Archives of Clinical Neuropsychology*, 20 (6), 745–753. [3]

WECHSLER, D. (1997): "WMS-III: Wechsler Memory Scale," in *Wechsler Memory Scale* (Third Ed.). San Antonio, TX: Psychological Corporation. [2]

WEINGART, L. R. (1992): "Impact of Group Goals, Task Component Complexity, Effort, and Planning on Group Performance," *Journal of Applied Psychology*, 77 (5), 682–693. [1]

Co-editor Guido Imbens handled this manuscript.

Manuscript received 11 May, 2020; final version accepted 10 April, 2021; available online 2 June, 2021.